



Do 11-month-old French infants process articles?

Pierre Hallé, Catherine Durand, Bénédicte de Boysson-Bardies

► To cite this version:

Pierre Hallé, Catherine Durand, Bénédicte de Boysson-Bardies. Do 11-month-old French infants process articles?. 2007. halshs-00129603

HAL Id: halshs-00129603

<https://shs.hal.science/halshs-00129603>

Preprint submitted on 8 Feb 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Do 11-month-old French infants process articles?

Pierre A. Hallé,^{abc} Catherine Durand,^b and Bénédicte de Boysson-Bardies^b

^a Laboratoire de Phonétique et Phonologie, CNRS-Paris 3

^b Laboratoire de Psychologie et Neurosciences Cognitives, CNRS-Paris 5

^c Haskins Laboratories

Running head: *Articles at 11 months*

Address for correspondence:

P. A. Hallé

Laboratoire de Phonétique et Phonologie

19 rue des Bernardins

75005 Paris

Email: pierre.halle@univ-paris5.fr

ABSTRACT

The first part of this study examined (Parisian) French-learning 11-month-old infants' recognition of the six definite and indefinite French articles: *le, la, les, un, une, des*. The six articles were compared with pseudo articles in the context of disyllabic or monosyllabic nouns, using the Head-turn Preference Procedure. The pseudo articles were similar to real articles in terms of phonetic composition and phonotactic probability, and real and pseudo noun phrases were alike in terms of overall prosodic contour. In three experiments, 11-month-old infants showed preference for real over pseudo articles, suggesting they have the articles' word-forms stored in long-term memory. The second part of the study evaluates several hypotheses about the role of articles in 11-month-olds infants' word recognition. Evidence from three experiments support the view that articles help infants to recognize the following words. We propose that 11-month-olds have the capacity to parse noun phrases into their constituents, which is consistent with the more general view that function words define a syntactic skeleton that serves as a basis for parsing spoken utterances. This proposition is compared to a competing account, which argues that 11-month-olds recognize noun-phrases as whole-words.

Keywords: determiners, recognition, parsing, noun phrases, early receptive lexicon

INTRODUCTION

The observation that children's early productions typically omit function words (Brown, 1973; Grégoire, 1937), has often led researchers to conclude that children first learn content words and only later begin to acquire function words. On this view, children would not form representations for function words before the age of two. This apparent difficulty has received two explanations. First, function words lack referential meaning whereas content words often have a concrete meaning with a real-world referent. Children thus might begin attending to content words and their referential value before they attend to function morphemes (Brown, 1973; MacNamara, 1982). A second reason for the difficulty associated with the acquisition of function words is their low perceptual salience and their variable acoustic forms.

The first explanation (lack of semantic motivation) can be undercut in several ways. As Naigles has argued (2002), "form is easy, meaning is hard." As an illustration, infants have an apparent advantage over toddlers in generalizing form categories from specific items because they simply build form representations whereas toddlers are engaged in the more demanding process of associating meanings to forms. More generally, as pointed out by Swingley (2005a), the logical developmental path in lexical acquisition is first to learn the sound-forms of frequently recurring words in the speech input, and later to begin "to figure out what those words mean." (Swingley, 2005a: 121). The acquisition of both function and content words could follow this path so that the child's initial stock of word-forms is not based on form-meaning associations but, rather, on frequently recurring forms in the input speech.

The second explanation focuses on the fact that function words are acoustically less salient than content words. This lack of salience could hinder the emergence of word-form representations for function words because children attend to and extract salient, stressed or final syllables more easily than non-salient, unstressed syllables (Echols & Newport, 1992). Both phonetically and phonologically, the spoken forms of function words are reduced

compared to those of content words in languages that are typologically as different as English, French, Chinese, and Turkish (Shi, 1995; Shi, Morgan, & Allopenna, 1998). One consequence of this difference is that newborns easily discriminate between function and content words (Shi, Werker, & Morgan, 1999). English-learning six-month-olds prefer content over function words (Shi & Werker, 2001), and the basis for this preference lies at the perceptual level (Shi & Werker, 2003). Function words may thus attract infants' attention to a lesser extent than content words because they are perceptually less "interesting," just like adult-oriented speech is less attractive than motherese (Fernald, 1985). But does this mean that infants are "deaf" to function words and cannot build word-form representations for them?

First, young infants can be trained to segment monosyllabic function words from spoken utterances, as shown in experiments using the "familiarization" version of the Head-turn Preference Procedure (HPP), in which infants are exposed to the target words in the training phase (Jusczyk & Aslin, 1995). The ability to segment function words was found for 7-to-9-month old German-learning infants (Höhle & Weissenborn, 2003), and for 6-to-8-month old Canadian French-learning infants months (Shi & Gauthier, 2005; Shi, Marquis, & Gauthier, 2006c), which is close to 7.5 months, the age at which English-learning infants can segment stressed monosyllabic words from continuous speech (Jusczyk & Aslin, 1995) and younger than 10.5 months, the age at which English-learning infants can be trained to retain most of the phonetic detail of unstressed syllables (Johnson, 2005). This evidence suggests that infants aged 8 months or younger can segment trained function words from continuous speech and thus have the capacity to build word-form representations for function words.

The next question that arises is whether children do build word-form representations for function words from natural exposure to the ambient speech. The early work of Gerken and colleagues showed that two-year-olds are sensitive to the correct use of functors in simple sentences (Gerken & McIntosh, 1993; Gerken, Landau, & Remez, 1990). Such sensitivity has now been found in 11-month-olds, using behavioral auditory preference tasks (Shady, 1996)

and ERP measurements (Shafer, Shucard, Shucard, & Gerken, 1998). Shady (1996) showed that 10.5-month-olds prefer listening to sentences with real rather than non-English functors. Shafer et al. (1998) found stronger ERP responses to tones superimposed on sentences with real as opposed to nonsense functors. More recently, Shi, Werker, and Cutler (2006a) directly addressed the issue of untrained determiner recognition by English-learning infants between the ages of 8 and 13 months, by comparing their responses to determiner+noun noun-phrases (NPs) consisting of a pseudo-noun (*breek* or *tink*) preceded either by a real determiner (*the*, *his*, *her*, *their*, or *its*) or a phonetically similar pseudo determiner (*kuh*, *ris*, *ler*, *lier*, or *ots*). Whereas no sign of preference for real over pseudo determiners is found at 8 months, a marginally significant preference appears at 11 months and becomes robustly significant at 13 months. Thus, phonetically detailed word-form representations for determiners would emerge at about 11 months for English-learning infants.

Whether French-learning infants can spontaneously recognize function words has not yet been investigated. The existing data for (Canadian) French-learning infants bear on *trained* function word segmentation, in experiments using the familiarization version of the HPP (Shi & Gauthier, 2005; Shi et al., 2006c). In this paper, we examine the issue of spontaneous function word recognition by Parisian French-learning infants, focusing on the age of 11 months, which seems critical for English-learning infants. It is possible that Parisian French-learning infants learn word-forms later than English-learning infants: recent data suggest that infants exposed to European French make use of prosodic cues to the segmentation of the speech input less efficiently than infants learning English (or Canadian French), paying more attention instead to segmental cues (Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006). This preference could thus lead to the delayed development of segmentation abilities, as observed in Nazzi et al. (2006), and, perhaps, to the delayed development of word-form representations. Yet infants learning Parisian French have generally built an early receptive “lexicon” of familiar content word word-forms at 11 months (Hallé & de Boysson-Bardies,

1994), which is no later than English- or Dutch-learning infants (Swingley, 2005b; Thierry, Vihman, & Roberts, 2003; Vihman, Nakai, DePaolis, & Hallé, 2004). But retaining familiar word-forms of function words may be more difficult for children because their spoken forms are more reduced than those of content words. On the other hand, their high frequency of occurrence and the distributional evidence for their phonetic cohesiveness may perhaps compensate for their lack of prosodic and phonetic salience. In Swingley's (2005a) modeling of how word-forms emerge from the input as distributionally cohesive speech chunks, functors, and in particular determiners, rank highly in a hierarchical clustering analysis, presumably because even though the phonetic context of any given determiner is highly variable, its own sequence of phonemes is always highly cohesive.

The first goal of the present study thus was to test the prediction that at least some functors' word-forms have become familiar to French-learning 11-month-old infants. More specifically, we examined the six definite and indefinite articles of French (*le, la, les, un, une, des*). We chose these articles because determiners in general are highly frequent function words and, amongst determiners, articles are about four times as frequent as the others.¹

The second goal of this study was to examine the role played by articles in word recognition. It seems that articles, or more broadly, function words, help listeners to parse the speech input. Research with artificial languages showed that "markers" (the equivalent of real language functors in that they have a low type-frequency but a high token-frequency) help participants to learn the "content words" of an artificial language (Valian & Coulson, 1988; Cutler, 1993). Cutler (1993) showed that learnability increases when markers (1) are frequent, and (2) are unstressed relative to non-markers, hence are closer equivalent of real-life functors. Valian and Coulson (1988) proposed that markers function as "anchor points" defining the grammatical skeleton of a utterance. Experimental data with children point to the same conclusion. Studies using "act-out" tasks have shown that two-years-olds respond better to commands issued using sentences with real as opposed to incorrect determiners or no

determiners at all (Gerken & McIntosh, 1993; Petetric & Tweney, 1977). These findings have been extended to 18-month-olds, using an on-line procedure that tracks the child participants' gaze (Kedar, Casasola, & Lust, 2006; Zangl & Fernald, 2003). In a recent paper, Fernald and Hurtado (2006) showed that 18-month-olds respond faster to object names in sentence frames than in isolation, and, interestingly, faster when the sentence frame contains a determiner than when it does not (e.g., *Look at the doggy!* vs. *Look. Doggy!*). Repetition task data provide another source of evidence. Gerken et al. (1990) found that 26-month-olds omit real functors but not nonsense functors when asked to repeat short sentences. Gerken proposed that they analyze real functors as morphemes whereas they treat nonsense functors as the beginning of the next word. The presence of a determiner would thus help children to parse the sentence and extract words correctly, thus helping them to learn new words. Recent data on Canadian French-learning eight-month-olds support that view. They learned trained words more easily when they appeared during familiarization in the context of a real rather than a pseudo determiner (Shi & Lepage, 2005; Shi, Cutler, Werker, & Cruickshank, 2006b). Similar findings have been reported for German-learning 11-month-olds (Höhle & Weissenborn, 2000): children learn *Kahn* ('boat') if trained with *der Kahn* ('the boat') not with *Vulkan* ('volcano'). Note that, here, the non-morphemic *Vul* is indeed the beginning of a word. These data suggest that infants as young as eight months of age do not treat determiners as possible word beginnings, although there might be cross-linguistic differences in the age at which this ability emerges, as suggested by the comparison of the Canadian French and the German data.

In this study, we ask whether Parisian French-learning 11-month-olds recognize the word-forms of articles and whether they already have the capacity to parse article+noun NPs.

Throughout the study, we use the HPP (Kemler Nelson, Jusczyk, Mandel, Myers, Turk, & Gerken, 1995) without familiarization. In the first part, we test whether 11-month-olds recognize untrained articles. In the second, we test whether articles help them to parse NPs.

EXPERIMENT 1

Do 11-month-old infants spontaneously recognize articles in the context of unfamiliar disyllabic nouns? The six articles of French are compared to pseudo articles matched to the real ones both phonetically and prosodically. A preference for the noun phrases with real articles would show that infants indeed know and recognize these articles.

Method

Participants. Sixteen 11-month-olds –eight girls and eight boys– from monolingual Parisian French-speaking families were tested. Their average age was 11 months and 3 days (range: 10.18-11.14 months.days, *SD*: 9.3 days). None of them was reported to suffer from audition deficits or temporary otitis. One additional boy was run but his results were eliminated because, on the post-test interview, he turned out to be exposed to Malagasy on a daily basis.

Stimuli and Design. Children were presented with two types of lists of 12 article+noun NPs (Appendix A) that contrasted real articles (the six non-partitive, not-elided, articles of French: *un, une, des, le, la, les*) with their paired pseudo-articles (/ɛr/, /ɔr/, /rœ/, /re/, /mã/, /kœ/). The nouns following the articles were the 12 unfamiliar disyllabic nouns used in Hallé and de Boysson-Bardies (1996). According to the “Lexique 3” French lexical database, their mean usage frequency was 2.9 occurrences per million (“film subtitles” database,² New, 2006). The pseudo-articles were roughly matched with the real ones in terms of phonetic composition and phonotactic probability. (Based on Tubach and Boë’s [1990] frequency counts, the two-phoneme sequence frequencies were 0.89 and 0.75 for the real and pseudo articles, respectively). The real articles agreed in terms of gender with the nouns (number agreement is not phonetically marked in French). One noun beginning with a vowel, *enzyme* /ãzim/, was paired with either *une* (/yn/) or /ɔr/ to avoid vowel-contact hiatus. A female speaker of

Parisian French, recorded the materials. She was instructed to produce the items in a natural, *non-motherese* style, at a comfortable rate, with an even tempo and intensity. To ensure prosodic homogeneity across the two types of NPs, the speaker read as a pair the two NPs for each noun (with a real, then a pseudo article). The speech was digitized at 16 kHz (16 bits, mono) and transferred to computer files. The pseudo and real article NPs were roughly similar with respect to overall prosodic shape in terms of pitch, intensity, or duration (see Appendix B). It was important to ensure prosodic similarity across the two types of NPs so that any observed preference on the part of the children could not be attributed to more salient prosody in some speech stimuli.³ Six pseudo-randomizations were prepared for each type of list, with the constraint that two consecutive NPs could not share the same article. Any two lists of the same type differed in their first two items, in order to ensure that infants had the opportunity to hear all the nouns as well as all the articles. Within each list, the items (mean duration 1.1 s) were presented approximately every 1.5 s. This design was used throughout the whole study.

Apparatus. The experiments reported below were conducted in a quiet, dimly lit room, kept at a comfortable temperature. Infants were seated on their parent's lap in the center of a three-sided booth, eyes at about 75 cm from the front panel. A small red lamp was mounted in the center of the front panel at eye level. A movie-camera was mounted behind this panel, just above the red lamp. A small blue lamp and a loudspeaker were mounted on each side panel, at eye level, at about 75 degrees from the center. The experimenter, unaware of the experimental settings and deaf to the stimuli, observed the infant on a video monitor (which relied on a low-luminance-sensitive camera) and measured on-line the infant's head turns by pressing or releasing the appropriate left or right button of a button box. A third "start" button was used to launch the trials. All the sessions were video taped, and later converted to digital movie files (Apple iMovie format) for off-line reliability evaluation. A PC micro-computer controlled the

audio stimulus presentation, the lightening of the various lamps, and recorded the on-line measurements of looking times.

Procedure. The procedure followed is common to all the experiments reported in this study. We followed the procedure used in Hallé and de Boysson-Bardies (1996) but introduced a few changes. First, the side of presentation was counterbalanced within instead of between subjects. However, the type and side of presentation of the *first* list of words were still counterbalanced between subjects, yielding four conditions (2 sides x 2 types for trial #1 of both training and test phases). Second, each experimental session was limited to two phases: a training and a test phase. As in our previous design, the training phase was intended to acquaint infants with the contingency between gaze-orientation and stimulus presentation.

In the training and test phases, each trial began with the red center light blinking to attract the infant's gaze. Once the experimenter judged the infant's gaze to be directed to the center light, she/he launched a trial by pressing the "start" button. Once this button was pressed, the computer turned off the center light and turned on the side light for the next auditory presentation. The side light kept on blinking for 2 s, then was turned off. Auditory presentation was initiated 0.5 s after the light started blinking. The infant was given 4 or 5 s (test or training phase) to begin orienting to the speech. If the child did not begin orienting within this time limit, the trial was abandoned (and the looking time scored as 0 s) and the computer proceeded to the next trial. When a trial was not eliminated in this way, the experimenter kept pressing the relevant button as long as the infant kept orienting to the speech. The experimenter would release or press the button again, depending upon the orientation of the infant's gaze. Looking away was allowed for no more than 2 or 3 s (test or training phase). If this time-out was exceeded, the current trial was abandoned and the computer proceeded to the next trial. Whenever a trial was abandoned, the audio presentation was terminated only after the current item had been entirely played. For each trial, the computer recorded the

cumulative orientation times. The training phase consisted of four trials with alternating list types. The test phase consisted of 12 trials (six of each type) in pseudo-random order, so that no more than two trials of the same type occur in a row.

Our procedure differs from the standard HPP in some respects. The main difference is that audio presentation starts after a short delay once the side lamp is turned on, whether the infant orients to the lamp or not. In the standard procedure, audio presentation is launched only when the infant orients to the side lamp. As a consequence, average looking times are shorter in our setting, in part because some trials are scored as 0 s: those for which the infant did not orient to the side lamp before time-out. On the other hand, the attrition rate is somewhat lower with our procedure than with the standard one. For instance, the attrition rate was 6.7% in Vihman et al.'s (2004) study (using our procedure) and is typically 15 to 20% with the standard procedure (Kemler Nelson et al., 1995). Another difference (in the present study) is that the side lamps are used only to initiate orientation before the onset of audio presentation. They are rapidly turned off and do not encourage further orientation. This feature also contributes to shorter looking times.

Results and Discussion

The results for this experiment and the following ones all are tabulated in Table 1. The mean looking times per trial did not differ for the two types of NPs: 3.34 vs. 3.16 s for real vs. pseudo articles, respectively, $|t(15)| < 1$, $p = .70$. The “preference ratio” (looking time to a given type divided by total looking time) was proposed in previous studies (Hallé & de Boysson-Bardies, 1994) as a way of factoring out individual variation in total attention span. The mean preference ratio for real articles was 0.53, not significantly above the 0.5 no-preference level, $t(15) = 1.31$, $p = 0.21$. The “split-half” analysis, first proposed in Vihman et al. (2004), was developed in order to see whether a preference becomes established right away or after some delay (also see Swingley, 2005b). We applied this analysis to the data (Figure 1).

An analysis of variance with the within-subject factors Article (real vs. pseudo) and Half (first vs. second half) revealed a main effect of Half, $F(1, 15) = 5.55, p < .05$ (looking times decreased across the two halves of the session) and a significant Article \times Half interaction, $F(1, 15) = 6.43, p < .05$, reflecting a reversal in the preference pattern across the two halves of the session. Preference for real over pseudo articles was nearly significant in the second half $t(15) = 2.05, p = .058$. The girls' performance did not differ significantly from that of the boys (0.55 and 0.52 preference ratios for real articles, respectively, $|t(14)| < 1$).

————— Figure 1 and Table 1 about here —————

To summarize, the preference for real over pseudo articles, which is not significant *overall*, only emerges in the second half of the session. In other words, article recognition, as indexed by a preference for real over pseudo article NPs, is not clear-cut in the context of an unfamiliar disyllabic noun. This context might be phonetically too rich for infants to recognize the articles quickly. Experiments 2.a and 2.b were designed to make the article more salient relative to the noun, so that infants' article recognition might become more clearly observable.

EXPERIMENTS 2.a and 2.b

As in Experiment 1, real and pseudo articles were compared in the context of a noun, which should not, by itself, determine a preference for one article type over the other. Monosyllabic nouns were used to increase articles' relative phonetic salience within each NP. In Experiment 2.a, only rare monosyllabic nouns were used, whereas four familiar words were inserted in Experiment 2.b to test for the impact of familiarity: increased familiarity with the nouns perhaps would affect infants' article recognition.

Method

Participants. Sixteen 11-month-olds –eight girls and eight boys– from monolingual French-

speaking families were tested in each experiment. Their average age was 11 months and 3 days (range: 10.21-11.18 m.d, *SD*: 9 days) and 11 months and 2 days (range: 10.18-11.26 m.d, *SD*: 11.2 days) in Experiments 2.a and 2.b, respectively. None of the infants suffered from any hearing problems. Two additional children participated in Experiment 2b but their results were not retained because of excessive fussiness or parental interference.

Stimuli. The same real and pseudo articles as in Experiment 1 were used. Experiment 2.a used 12 rare nouns, whose mean frequency was 6 o.p.m. (“film subtitles” subpart of Lexique 3). None of them was phonetically close to an existing familiar word. Experiment 2.b used four of these 12 rare nouns (9 o.p.m.), four familiar words (181 o.p.m.) appearing in our corpus of utterances of the early productive lexicon (de Boysson-Bardies, 1996), and four nouns of intermediate usage frequency (117 o.p.m.) which are absent from that corpus. The lists of NPs are shown in Appendix A. The NPs were recorded by the same speaker as in Experiment 1, with the same instructions to ensure prosodic homogeneity across NP pairs (Appendix B).

Results and Discussion

In both experiments, looking times were significantly longer for real than for pseudo-articles (2.a: 4.5 s vs. 2.7 s per trial, $t(15) = 2.47$, $p < .05$; 2.b: 4.0 s vs. 2.8 s, $t(15) = 2.34$, $p < .05$). The preference ratios for real over pseudo articles also were significantly above 0.5 (2.a: 0.63, $t(15) = 3.09$, $p < .01$; 2.b: 0.57, $t(15) = 3.45$, $p < .005$). A large majority of infants preferred the real articles: 13 and 14 (out of 16) in Experiments 2.a and 2.b, respectively. Girls and boys did not differ in their performances, although the preference for real articles across the two experiments was numerically larger for boys than for girls (0.63 vs. 0.57).

Increased noun familiarity did not affect the detection of the preceding article: the preference for real over pseudo articles was equivalent in Experiments 2.a and 2.b, with a somewhat tighter distribution of preference ratios in Experiment 2.b.

In summary, the data of Experiments 1 and 2 showed that 11-month-olds can recognize articles. The representations they have formed for articles may not be fully specified phonetically, but they are sufficient to distinguish between pairs such as /re/-/lœ/, or /kœ/-/le/. Article recognition, however, is clearly observable only in short utterances with monosyllabic nouns. It is more difficult (it takes more time) in the context of unfamiliar disyllabic nouns.

We turn now to the issue of the role played by articles in the processing of disyllabic noun NPs by 11-month-olds infants. A first possibility is that articles make the recognition of a following noun more difficult. However, pilot work conducted in the mid '90s did not support this possibility.⁴ The evidence for word segmentation capacities in English-learning infants (Jusczyk, Houston, & Newsome, 1999) also argues against this possibility: 11-month-olds segment familiar disyllabic words from a utterance. However, *French*-learning infants might not exhibit this capacity prior to 16 months (Nazzi et al., 2006; but see Polka & Sundara, 2003). Infants might also recognize NPs with a familiar noun as whole-words they have been frequently exposed to. Yet, the recognition of a “familiar” NP might be less easily observable than that of a familiar noun because any article+noun NP is necessarily heard less often than the noun alone. For example, according to “Lexique 3” (“film subtitles”), the frequency of *gâteau* (‘cooky’) is 57.1 o.p.m., while that of the sequence *le gâteau* (‘the cooky’) is only 9.7.

A second possibility is that articles are simply ignored because of their low perceptual salience. On this view, 11-month-old infants would segment familiar nouns from article+noun NPs, regardless of whether articles are real or not. We call this hypothesis the “noun-segmentation” hypothesis. A variant of this view is that any NP is treated as a whole-word equivalent of the embedded noun, provided that its prosodic profile consists of an unstressed syllable followed by a stressed noun. This “prosodic whole-word” hypothesis also predicts that word recognition does not depend on whether articles are real articles.

Finally, articles could help children recognize the following noun if they parse the NP into article plus noun. By this “parsing” hypothesis, infants would show a marked preference for familiar noun NPs *if and only if* the familiar nouns are preceded by a real article.

Experiment 3 examines the first possibility listed above, that the presence of a real article makes familiar word recognition more difficult.

EXPERIMENT 3

We compare here NPs with a familiar noun to NPs with an unfamiliar noun, both types of NP beginning with a real article. If real articles do not impede familiar word recognition, preference ratios for NPs with a familiar noun should be similar to those for isolated familiar nouns reported in previous studies.

Method

Participants. Sixteen 11-month-old infants –eight girls and eight boys– from monolingual French-speaking families were tested. Their average age was 11 months and 4 days (range: 10.16-11.15 m.d, *SD*: 10.4 days). None of the infants suffered from any hearing problems. Two additional children were run but their data were not retained: one boy failed to orient to the speech for more than 1 s per trial; one girl rapidly became excessively fussy.

Stimuli. The six French articles were used. The 12 disyllabic familiar nouns and 12 disyllabic rare nouns in Hallé and de Boysson-Bardies (1996) were used (Appendix A). The NPs were recorded by the same speaker as in Experiments 1-2; NPs were paired by article to ensure prosodic homogeneity (Appendix B).

Results and Discussion

Looking times were significantly longer for NPs with familiar than unfamiliar nouns,

and the preference ratios for familiar over unfamiliar NPs were significantly above the 0.5 no-preference level (looking times: 3.7 s vs. 2.5 s per trial, $t(15) = 2.31$, $p < .05$; preference ratios: 0.61, $t(15) = 3.53$, $p < .005$). There was no significant difference between girls and boys (0.60 and 0.63 preference ratios, respectively).

Eleven-month-olds showed a clear preference for familiar over rare words following a real article. The preference ratios are not lower than those found for isolated nouns in Hallé and de Boysson-Bardies (1994): 0.61 in this experiment against 0.57 in the '94 study. Thus, real articles do not make word recognition more difficult. On the “noun-segmentation” hypothesis, infants segment familiar nouns from article+noun NPs, whether the articles are real or not. On the “prosodic whole-word” hypothesis, infants treat familiar NPs as variants of familiar nouns. Thus, on both hypotheses, replacing the real articles with pseudo articles, while maintaining the overall prosodic shape of the NPs, should not alter infants' preference for familiar over unfamiliar nouns found in Experiment 3. On the “parsing” hypothesis, pseudo articles would not allow infants to parse NPs into article plus noun, and no infant preference for familiar over unfamiliar words preceded by pseudo articles should be found.

EXPERIMENT 4

Experiment 4 compares familiar and unfamiliar noun NPs beginning with a pseudo article. Both the noun-segmentation and the prosodic whole-word hypotheses, but not the parsing hypothesis, predict that 11-month-olds can recognize familiar words in such context.

Method

Participants. Sixteen 11-month-olds –eight girls and eight boys– from monolingual French-speaking families were tested. Their average age was 10 months and 30 days (range: 10.18–11.15 m.d, SD : 10.2 days). None of the infants suffered from any hearing problems. Two

additional children were run but their data were eliminated: one boy cried and did not complete the session; one girl failed to orient to the speech for more than 1 s per trial.

Stimuli. The six pseudo articles of Experiments 1-2 and the nouns of Experiment 3 (12 familiar vs. 12 unfamiliar) were used (Appendix A). The NPs were recorded by the same speaker as in Experiments 1-3, in the same way as for Experiment 3.

Results and Discussion

Looking times for NPs with familiar nouns were not longer than those with unfamiliar nouns (2.8 s vs. 2.9 s per trial, $|t| < 1$), and the preference ratios for familiar over unfamiliar NPs did not differ from 0.5 (0.51, $|t| < 1$). There was no significant difference between girls and boys (0.49 and 0.53 preference ratios, respectively).

The results of Experiments 3 and 4 are thus clearly different,⁵ even though the pseudo and real articles used in these experiments were very similar in terms of their prosodic characteristics and rather close in terms of their phonetic composition. The lack of preference in Experiment 4 for familiar over unfamiliar NPs is clearly at odds with both the “noun-segmentation” and the “prosodic whole-word” hypotheses. But we cannot rule out a more specific “holistic noun phrase” account whereby infants would recognize only a restricted set of NPs, rather than any NP composed of an unstressed syllable plus a familiar noun. We return to this issue in the General Discussion but we note here that the phonetic differences between the pseudo and real articles induce dramatic processing differences: real articles allow for the recognition of a following familiar noun; pseudo articles preclude that recognition.

A straightforward account of the results, consistent with the parsing hypothesis, is proposed by Gerken (Gerken et al., 1990): infants analyze real articles “as morphemes but [treat] nonsense functors as part of the adjacent content words.” By this account, all the NPs used in Experiment 4 were treated as three-syllable unfamiliar words. Thus we found a lack of

preference for one type over the other. This conclusion, however, is based on a null result. It would be strengthened if supported by a positive result. Gerken's account predicts that *or chaussure* would be an unrecognized word-form while *la chaussure* would be analyzed as a determiner plus noun NP. Eleven-month-olds should thus recognize familiar words in the latter but not the former type of NP, showing a preference for *la chaussure* over *or chaussure*.

EXPERIMENT 5

In Experiment 5, as in Experiments 1 and 2, real articles are compared to pseudo articles. But the following nouns are disyllabic *familiar* nouns. The parsing hypothesis predicts that only real article NPs are parsed and their constituents recognized. For instance, *chaussure* should be recognized in *la chaussure*, not in *or chaussure*.

Method

Participants. Sixteen 11-month-olds –eight girls and eight boys– from monolingual French-speaking families were tested. Their average age was 11 months and 6 days (range: 10.23–11.18 m.d, *SD*: 8.8 days). None of the infants had any hearing problems. All were run successfully.

Stimuli. The same real and pseudo articles as in Experiments 1–2 were used. The nouns were the familiar words used in Experiment 3 (Appendix A). The NPs were recorded by the same speaker as in Experiments 1–4, in the same way as for Experiments 1–2.

Results and Discussion

Looking times were significantly longer for NPs with real rather than pseudo-articles, and the preference ratios for real over pseudo articles were significantly above 0.5 (looking

times: 3.5 s vs. 2.4 s per trial, $t(15) = 3.49$, $p < .005$; preference ratios: 0.59, $t(15) = 4.10$, $p < .001$). Girls did not differ from boys (0.57 and 0.60 preference ratios, respectively).

The data suggest that 11-month-olds can segment and recognize familiar words embedded in article+noun NPs only for NPs with a real article. The finding of Experiment 4, that infants do not recognize *or chaussure* as a more familiar form than *or berline*, is fully consistent with the preference found here for *la chaussure* over *or chaussure*. The two results combine to demonstrate that even though *or chaussure* contains a familiar noun, the presence of the pseudo article *or* makes that noun difficult for infants to recognize. The clear preference for *la chaussure* over *or chaussure* thus cannot be explained by the sole preference for *la* over *or*, which we found in Experiments 2a-b and, to a lesser degree, in Experiment 1. It must be mainly due to the fact that infants recognize familiar words after a real article (or no article) but not after a pseudo article. We return to this point in the General Discussion.

Taken together, the results of Experiments 4 and 5 do not support the two accounts discussed earlier, the noun-segmentation and the prosodic whole-word accounts, which both predicted preferences determined by noun familiarity alone. Rather, they are consistent with a parsing account: real articles trigger a parse, which may produce a known noun in addition to the article. The possibility that familiar-noun NPs may be recognized holistically as whole-words for a restricted set of determiners is discussed in the General Discussion.

GENERAL DISCUSSION

The first part of this study showed that (Parisian) French-learning 11-month-old infants can segment untrained real articles from NPs. That is, infants have already stored in memory the word-forms of these articles and can detect them in simple NPs. Article segmentation from NPs, however, is not extremely robust in the sense that detection seems more difficult when articles precede unfamiliar disyllabic rather than monosyllabic nouns (Experiments 1 and 2.a).

The preference for real over pseudo article NPs was replicated in Experiment 2.b, which included four presumably familiar nouns.

Experiment 3 showed that real articles are not detrimental to disyllabic word recognition. Rather, the results suggest that real articles allow infants to recognize a following noun (Experiments 3 and 5) whereas pseudo articles, however close to real articles they may be in terms of prosodic and phonetic shape, completely block that recognition (Experiment 4). As a straightforward and parsimonious account of these data, we propose that 11-month-old infants successfully parse NPs consisting of a real article plus a familiar noun.

The parsing account readily explains why the infant preference found for real over pseudo article NPs was much less clear-cut in Experiment 1 (disyllabic unfamiliar nouns) than in Experiment 5 (disyllabic familiar nouns). Indeed, in the former case, the attempted parsing of NPs produced a known article plus an unidentified portion; in the latter case, the parsing was successful in that it resulted in two known items. In other words, the preference infants demonstrated in Experiment 5 reflects something more than successful segmentation. It is successful parsing into known elements. We did not observe the same difference between Experiments 2.a and 2.b, in which monosyllabic nouns were used. In both experiments, we found a clear preference for real over pseudo articles, in an “easier” context than that of disyllabic unfamiliar nouns. It is possible that infants parsed some NPs into article plus noun in Experiment 2.b (e.g., *un pied*) but not enough to produce an observable increase in preference as compared to Experiment 2.a, in which that preference was already quite strong. In contrast, there was much more leeway for an increase in preference to be observed when unfamiliar nouns were substituted for familiar nouns in a disyllabic context. Our proposition that 11-month-olds parse noun phrases such as *la chaussure* is therefore based on the disyllabic noun data. Additional research, using a more sensitive paradigm, is needed in order to determine whether parsing also takes place with monosyllabic nouns.

The kind of parsing we propose presumably occurs at the word-form level. That is, French 11-month-olds only “know” that the word-forms *le* or *la* are possibly followed by a familiar word-form. Such knowledge could be a first step toward mapping word forms to presumably innate categories (see Valian, in press). It is unlikely, though, that infants that young can use articles for a full morphosyntactic analysis and assign a precise grammatical category to a word. The capacity to distinguish, for example, determiners from personal pronouns does not seem to appear before 14-16 months for German-learning children (Höhle, Weissenborn, Kiefer, Schulz, & Schmidt, 2004). Likewise, Shady (1996) found that English-learning infants younger than 16 months are insensitive to the interchanging of function words in made-up passages. That is, 11-month-olds, for example, accept equally well “...had seen that bike...” and “...that seen had bike....” These findings suggest that French-learning 11-month-olds’ detection of articles does not help other than possibly signaling an adjacent familiar word. On the other hand, not detecting an article in a noun phrase precludes recognition of the noun. To sum up, our account is actually quite consistent with Gerken et al.’s (1990) proposition: function morphemes are analyzed as such whereas nonsense functors are treated as part of the adjacent word. In other words, NPs can be parsed into constituents if and only if they begin with a real article.

Parsing noun phrases only at the level of word-forms seems a modest achievement. Yet, this ability gives infants a head start on lexical acquisition. Indeed, infants may attend to what is produced after any article they have detected in a utterance. In that sense, article detection could help infants to learn new word-forms, in line with the more general claim that function morphemes play an important role in lexical bootstrapping (Christophe, Guasti, Nespor, Dupoux, & van Ooyen, 1997; Gout, Christophe, & Morgan, 2004).

With respect to real article detection, it might be argued that infants simply detect frequent *syllables*, not familiar words (Nazzi et al., 2006). This claim is not supported by the radically different outcomes of Experiment 1 and 5, which both compared the same real

articles to pseudo articles in disyllabic contexts. A more interesting question is what may support article detection. One obvious factor is the frequency of occurrence of the form in the speech infants are exposed to. According to lexical counts of spoken French (New, 2006), all six articles are highly frequent (footnote 1). (But the frequencies may differ in the speech addressed to infants.) Future research could address the issue of frequency by using a different paradigm or a different design comparing frequent article(s) and less frequent determiner(s) (cf. Shi et al., 2006b). Another issue is how detailed infants' representations of the articles are. Again, these questions require future research.

We return finally to the account that takes the results of Experiment 3-5 as showing that real article+noun NPs are recognized as whole-words. For example, *le gâteau* ('the cooky') could have been learned as a whole and preferred over *le défaut* ('the defect') just like *gâteau* is preferred over *défaut* (Hallé & de Boysson-Bardies, 1994). We have distinguished two extreme versions of the whole-word account. On a "loose" prosodic whole-word account, any weak syllable followed by a familiar word is treated as an acceptable variant of the familiar word. As discussed earlier, the outcome of Experiments 4-5 did not support this account: *ré gâteau* was not as acceptable as *le gâteau* and treated as a variant of the familiar word *gâteau*. At the other extreme, NPs could be stored as whole-words in all their phonetic detail. That is, most of the NPs consisting of a real article followed by *gâteau* could be familiar word-forms for 11-month-olds. Because form-meaning associations are probably not firmly established at this age, or established at all (Swingley, 2005a: 121), *gâteau* and the derived NPs possibly familiar to 11-month-olds would not necessarily be represented as related word-forms. If they were, that is, if *le gâteau* was recognized by infants as equivalent to *gâteau*, this whole-word account would be functionally indiscernible from the parsing account we propose. If they were not, the whole-word account would not be compelling for at least two reasons. First, storing many related word-forms without recording their commonality would be uneconomical in that

infants would have to postpone the work of sorting out word-forms to a later stage when they discover form-meaning associations. Second, this whole-word scenario entails that infants lack the capacity to combine the building blocks they have extracted from the input speech, a capacity which may be viewed at the heart of language acquisition and use. On a more empirical note, given that infants identify and retain word-forms on the basis of how cohesive they are in terms of co-occurring sounds (cf. Saffran, Newport, & Aslin, 1996; Swingley, 2005a; the idea dates back to Harris, 1955), speech chunks such as *gâteau* and *le gâteau* might indeed be both retained as possible word-forms in a “protolexicon.” Yet, *le gâteau* is obviously much less cohesive than *gâteau*, leading to the prediction that it is less firmly established in 11-month-olds’ protolexicon, thus should be less easily recognized. This prediction is not supported by the present data: the preference for *le gâteau* over *le défaut* (Expt. 3) is as robust as the preference found for *gâteau* over *défaut* in previous studies. The equivalent results for familiar nouns and familiar NPs thus suggest that familiar NPs are parsed by 11-month-olds, producing, in particular, known familiar nouns. Articles thus appear to function as landmarks which potentially help children to discover novel word-forms.

Children, however, may fail to correctly extract word-forms in certain situations. This can be the case for words beginning with a vowel. Empirical data showed that infants do not segment familiarized vowel-initial words from continuous speech before they are 16 month old (Mattys & Jusczyk, 2001). Pye (1983) observed that children learning a language in which vowel-initial words can be resyllabified frequently produce segmentation errors. In French, vowel-initial words are resyllabified so that their initial vowel becomes part of a CV syllable whose C belongs to the previous word (often an article) in enchaînement, elision, and liaison situations (e.g., *une auto* /y.noto/ ‘a car’, *l’oiseau* /lwazo/ ‘the bird’, *un ours* /ẽ.nurs/ ‘a bear’). In these cases, in order to extract the correct word-form (e.g., *auto*, *oiseau*, *ours*), children have to go beyond a syllable-based distributional analysis. They have to detect co-occurrences at

the phonemic level, regardless of syllabic structure. The current evidence suggests that early word-forms are extracted from input speech due to a bias to cluster *syllables* that often co-occur (Goodsitt, Morgan, & Kuhl, 1993; Saffran et al., 1996; Swingley, 2005a). Statistical clustering of phonemes producing word-forms that are misaligned with syllable boundaries is probably much more difficult. This is suggested by child production errors reflecting mis-segmentation at the phoneme level (e.g., **papa nours* instead of *papa ours*, Chevrot & Fayol, 2001) as well as by diachronic changes reflecting article agglutination (e.g., French *lierre* ‘ivy’ < Latin *hedora*; French *lendemain* ‘following day’ < *l’en demain*). So, while articles probably help children to discover novel word forms most of the time by providing a clear landmark, the mechanism is not fool-proof and, on occasion, children might store incorrect word-forms in their protolexicon.

Author Note

We are grateful to all the parents who graciously spent their time bringing their babies to our lab. We would also like to thank Claude Kervella, who designed the entire experimental setting, and Souhila Messaoud-Galusi, who helped run the children. We also thank the editor, Elizabeth Johnson, one anonymous reviewer, and Andrea Levitt, whose comments greatly helped us to improve an early version of the manuscript

Footnotes

¹ Frequencies of occurrence (o.p.m.) according to the “Lexique 3” database (New, 2006; also see New, Pallier, Brysbaert, & Ferrand, 2004): from 13,987 to 5,748 (average 10,162) for *la, le, un, les, une, des* in that order, against an average 2,608 for the other determiners (including the partitive articles *du* and *de la*). The definite and indefinite articles of French are thus clearly the most frequent function words, all above the five-per-thousand frequency level.

² The “Lexique 3” database is subdivided into a “film subtitles” and a “printed text” subpart, each based on a corpus of about 16 million words. The “film subtitles” subpart is presumably the more representative of spoken French. It is not, however, child-directed speech.

³ Although pseudo article did not differ from real article in terms of duration and intensity, they happened to have somewhat more marked F0 contours in this experiment.

⁴ In this unpublished study, 10 French 11-month-olds were tested on their preference for familiar over unfamiliar nouns preceded by real articles. Nine infants preferred the familiar nouns, as indicated by the longer looking times for noun phrases with familiar (6.1 s) than unfamiliar nouns (3.9 s), $t(9) = 3.14$, $p < .05$. The preference ratio (for familiar nouns) was 0.59. Unfortunately, the study could not be completed due to logistical reasons.

⁵ Statistic analyses were run to compare the outcomes of Experiments 3 and 4, with Article (real in Expt. 3 vs. pseudo in Expt. 4) and Noun (familiar vs. unfamiliar) as between- and within-subject factors, respectively. The Article \times Noun interaction was nearly significant, $F(1, 30) = 3.79$, $p = .061$, and indeed, Noun was significant for real articles, $F(1, 15) = 5.35$, $p < .05$, but not for pseudo articles, $F < 1$. Consistent with this difference, the preference ratios for familiar over unfamiliar nouns were higher for real than pseudo articles (0.61 vs. 0.51), $F(1, 30) = 4.35$, $p < .05$.

REFERENCES

- BROWN, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.
- CHEVROT, J.-P. & FAYOL, M. (2001). Acquisition of French liaison and related child errors. In *Proceedings of the 8th Conference of the International Association for the Study of Child Language* (pp. 760-774). Somerville, MA: Cascadilla Press.
- CHRISTOPHE, A., GUASTI, M. T., NESPOR, M., DUPOUX, E., & van OOOYEN, B. (1997). Reflections on prosodic bootstrapping: its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, **12**, 585-612.
- CUTLER, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research*, **22**, 109-131.
- de BOYSSON-BARDIES, B. (1996). *Comment la parole vient-elle aux enfants ?* [How language comes to children?]. Paris: Odile Jacob.
- ECHOLS, C., & NEWPORT, E. (1992). The role of stress and position in determining first words. *Language Acquisition*, **2**, 189-220.
- FERNALD, A. (1985). Four-month-olds prefer to listen to motherese. *Infant Behavior and Development*, **8**, 181-195.
- FERNALD, A., & HURTADO, N. (2006). Names in frames: Infants interpret words in frames faster than words in isolation. *Developmental Science*, **9**, F33-F40.
- GERKEN, L. & McINTOSH, B. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology*, **29**, 448-457.
- GERKEN, L.A., LANDAU, B., & REMEZ, R. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology*, **27**, 204-216.
- GOODSITT, J. V., MORGAN, J. L., & KUHL, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, **20**, 229-252.

- GOUT, A., CHRISTOPHE, A., & MORGAN, J. (2004). Phonological phrase boundaries constrain lexical access: II. infant data. *Journal of Memory and Language*, **51**, 548–567.
- GRÉGOIRE, A. (1937). *L'apprentissage du langage* [Language apprenticeship]. Paris: E. Droz.
- HALLÉ, P., & de BOYSSON-BARDIES, B. (1994). Emergence of an early lexicon: Infant's recognition of words. *Infant Behavior and Development*, **17**, 119-129.
- HALLÉ, P., & de BOYSSON-BARDIES, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, **19**, 465-483.
- HARRIS, Z. S. (1955). From phoneme to morpheme. *Language*, **31**, 190–222
- HÖHLE, B., & WEISSENBORN, J. (2000). The origins of syntactic knowledge: Recognition of determiners in one year old German children. In *Proceedings of the 24th Annual Boston Conference on Language Development* (pp. 418-429). Somerville, MA: Cascadia Press.
- HÖHLE, B., & WEISSENBORN, J. (2003). German learning infants' ability to detect unstressed closed class elements in continuous speech. *Developmental Science*, **6**, 122-127.
- HÖHLE, B., WEISSENBORN, J., KIEFER, D., SCHULZ, A., & SCHMIDT, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, **5**, 341-353.
- JOHNSON, E. (2005). English-learning infants' representations of word-forms with iambic stress. *Infancy*, **7**, 99-109.
- JUSCZYK, P., & ASLIN, R. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, **29**, 1-23.
- JUSCZYK, P., HOUSTON, D., & NEWSOME, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, **39**, 159-207.
- KEDAR, Y., CASASOLA, M., & LUST, B. (2006). Getting There Faster: 18- and 24-Month-Old Infants' Use of Function Words to Determine Reference. *Child Development*, **77**, 325-

338.

KEMLER NELSON, D., JUSCZYK, P., MANDEL, D., MYERS, J., TURK, A., & GERKEN, L. (1995). The Head-Turn Preference Procedure for testing auditory perception. *Infant Behavior and Development*, **18**, 111-116.

MacNAMARA, M. (1982). *Names for things*. Cambridge, MA: MIT Press.

MATTYS, S., & JUSCZYK, P. (2001). Do infants segment words or recurrent contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, **27**, 644-655.

NAIGLES, L. (2002). Form is easy, meaning is hard: resolving a paradox in early child language. *Cognition*, **86**, 157-199.

NAZZI, T., IAKIMOVA, G., BERTONCINI, J., FRÉDONIE, S., & ALCANTARA, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, **54**, 283-299.

NEW, B. (2006, April). Lexique 3: une nouvelle base de données lexicales. Communication presented at the TALN 2006 meeting (Traitement Automatique des Langues Naturelles). Louvain, Belgique.

NEW, B., PALLIER, C., BRYSSBAERT, M., & FERRAND, L. (2004) Lexique 2 : A New French Lexical Database. *Behavior Research Methods, Instruments, & Computers*, **36**, 516-524.

PETRETIC, P., & TWENEY, R. (1977). Does comprehension precede production? The development of children's responses to telegraphic sentences of varying grammatical adequacy. *Journal of Child Language*, **4**, 56-73.

PYE, C. (1983). Mayan Telegraphese: Intonational determinants of inflectional development in Quiche Mayan. *Language*, **59**, 583-604.

- POLKA, L., & SUNDARA, M. (2003). Word segmentation in monolingual and bilingual infant learners of English and French. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1021–1024). Barcelona, Spain.
- SAFFRAN, J., ASLIN, R., & NEWPORT, E. (1996). Statistical learning by 8-month-old children. *Science*, **274**, 1926-28.
- SHADY, M. (1996). *Infants' sensitivity to function morphemes*. Doctoral dissertation, the State University of New York: University at Buffalo.
- SHAFFER, V., SHUCARD, D., SHUCARD, J., & GERKEN, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language, and Hearing Research*, **41**, 874-886.
- SHI, R. (1995). *Perceptual correlates of content words and function words in early language input*. Doctoral dissertation, Brown University.
- SHI, R., CUTLER, A., WERKER, J., & CRUICKSHANK, M. (2006b). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of the Acoustical Society of America*, **119**, EL61-EL66.
- SHI, R., & GAUTHIER, B. (2005). Recognition of function words in 8-month-old French-learning infants. *Journal of the Acoustical Society of America*, **117**, 2426 [abstract].
- SHI, R. & LEPAGE, M. (2005, July). Function words in preverbal infants' emerging lexicon. Communication presented at the CSBBCS (Canadian Society for Brain, Behaviour and Cognitive Science) meeting. Université de Montréal, Canada.
- SHI, R. MARQUIS, A., & GAUTHIER, B. (2006c). Segmentation and representation of function words in preverbal French-learning infants. In *Proceedings of the 30th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
- SHI, R., MORGAN, J. L., & ALLOPENNA, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*, **25**, 169-201.

SHI, R., & WERKER, J. (2001). Six-month-old infants' preference for lexical words.

Psychological Science, **12**, 70-75.

SHI, R., & WERKER, J. (2003). The basis of preference for lexical words in 6-month-old infants. *Developmental Science*, **6**, 484-488.

SHI, R., WERKER, J., & CUTLER, A. (2006a). Recognition and representation of function words in English-learning infants. *Infancy*, **10**, 187-198.

SHI, R., WERKER, J., & MORGAN, J. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, **72**, B11-B21.

SWINGLEY, D. (2005a). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, **50**, 86-132.

SWINGLEY, D. (2005b). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, **8**, 432-443.

THIERRY, G., VIHMAN, M., & ROBERTS, M. (2003). Familiar words capture the attention of 11-month olds in less than 250 ms. *NeuroReport*, **18**, 2307-2310.

TUBACH, J.-P., & BOË, L.-J. (1990). *Un corpus de transcription phonétique (300,000 phones)* (Paris France Télécom Internal Report 90 D 002). Paris: France Télécom.

VALIAN, V. (in press). Abstract linguistic representations and innateness: The development of Determiners. In W. Lewis, S. Karimi, H. Harley, & S. Farrar (Eds.), *Time and again* [tentative title]. Amsterdam: John Benjamins.

VALIAN, V. & COULSON, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, **27**, 71-86.

VIHMAN, M., NAKAI, S., DePAOLIS, R., HALLÉ, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, **50**, 336-353.

ZANGL, R., & FERNALD, A. (2003, October). Sensitivity to function word morphemes in online sentence processing: Developmental changes from 18 to 36 months. Talk presented at the Boston University Conference on Language Development, Boston, MA.

Appendix A. Speech materials in Experiments 1-5, labeled E1 to E5. In Experiments 3-4, the compared NPs differ by the noun, while they differ by the article in the other experiments.

article		E1	E2.a	E2.b	E5
real	pseudo	unfamiliar	unfamiliar	mixed	familiar
des	roe	bigots	mèches	moines	canards
des	roe	tangages	boucs	pommes	lapins
la	man	berline	cure	cure	chaussure
la	man	volute	natte	fleur	voiture
le	ré	caduc	dôme	dôme	bonjour
le	ré	défaut	lard	ciel	gâteau
les	koe	félins	dunes	mains	ballons
les	koe	soudards	pions	trains	chapeaux
un	èr	busard	moine	chat	biberon
un	èr	cobaye	soc	pied	oiseau
une	or	enzyme	tuile	tuile	banane
une	or	licence	craie	balle	poupée

article		noun	
E3	E4	familiar	unfamiliar
real	pseudo		
des	roe	canards	bigots
des	roe	lapins	tangages
la	man	chaussure	berline
la	man	voiture	volute
le	ré	bonjour	caduc
le	ré	gâteau	défaut
les	koe	ballons	félins
les	koe	chapeaux	soudards
un	èr	biberon	busard
un	èr	oiseau	cobaye
une	or	banane	enzyme
une	or	poupée	licence

Appendix B. Prosodic characteristics of the NPs used in Experiments 1-5. The first column gives an example of NP, the second is the percent duration of the article in the NP; the other columns indicate the duration (ms), mean F0, F0 standard deviation and F0 range in Hz, and max intensity (dB) for the NP's article and noun. Significant differences between the two types of NP compared in any given experiment are indicated with '*'.

examples	%dur.	dur.	article				dur.	noun			
			mF0	sdF0	rgeF0	int.		mF0	sdF0	rgeF0	int.
E1											
le caduc	18.2	155	156	6.8	25	77.2	699	179	49.5	193	77.9
ré caduc	19.7	175	160	12.3	41	76.9	715	176	47.5	184	76.4
E2.a											
des mèches	23.2	190	168	8.8	33	76.4	629	186	45.3	157	75.0
rœ mèches	25.2	208	170	11.4	39	76.8	617	187	46.8	146	74.6
E2.b											
les trains	22.4	169	188	17.0	49	78.5	584	157	26.1	98	77.9
kœ trains	24.1	188	190	19.9	59	79.9 *	593	154	25.2	101	76.8 *
E3											
des canards	18.1	162	146	5.7	22	73.0	734	180	40.0	149	74.6
des bigots	18.2	173	142	7.0	30	72.7	780	182	40.2	148	74.7
E4											
èr biberon	21.1	185	146	8.9	35	75.1	692	183	42.5	157	76.3
èr busard	20.4	186	142	9.2	38	75.3	727	180	39.8	144	77.7
E5											
une poupée	19.3	159	146	5.8	22	74.9	665	168	42.2	134	76.7
or poupée	19.0	157	148	8.8	32	75.4	667	165	40.1	129	76.6

Table 1. Mean looking times (s) and preference ratio in 1st, in 2nd half of test, and overall, for Experiments 1-5 (E1-5). Ratios significantly above/below 0.5 are signaled with ‘*’

		<u>Type of NP</u>		<u>ratio: A/(A+B)</u>
		<u>A</u>	<u>B</u>	
E1		le soudard	ré soudard	
	1st half	3.6	4.4	0.47
	2nd half	3.1	1.9	0.62 *
	<u>overall</u>	<u>3.3</u>	<u>3.2</u>	<u>0.54</u>
E2.a		le dôme	ré dôme	
	1st half	5.3	3.1	0.63 *
	2nd half	3.6	2.2	0.63 *
	<u>overall</u>	<u>4.5</u>	<u>2.7</u>	<u>0.63 *</u>
E2.b		le chat	ré chat	
	1st half	5.3	3.6	0.60 *
	2nd half	2.8	2.0	0.56 *
	<u>overall</u>	<u>4.0</u>	<u>2.8</u>	<u>0.58 *</u>
E3		le canard	le soudard	
	1st half	4.0	3.0	0.57 *
	2nd half	3.4	2.0	0.63 *
	<u>overall</u>	<u>3.7</u>	<u>2.5</u>	<u>0.60 *</u>
E4		ré canard	ré soudard	
	1st half	3.0	2.8	0.52
	2nd half	2.7	3.0	0.49
	<u>overall</u>	<u>2.8</u>	<u>2.9</u>	<u>0.50</u>
E5		le canard	ré canard	
	1st half	4.3	2.4	0.67 *
	2nd half	2.6	2.4	0.49
	<u>overall</u>	<u>3.5</u>	<u>2.4</u>	<u>0.58 *</u>

Figure caption

Figure 1. Mean looking times per trial (sec.) for the real and pseudo article NPs in the first and second half of the test (Experiment 1).

Figure 1

